Mapping urban deprivation from Sentinel 1/2 using artificial intelligence and weakly labelled data

Julien Govoorts
ANAGEO - DGES
Université Libre de Bruxelles
Brussels, Belgium

Taïs Grippa
ANAGEO - DGES
Université Libre de Bruxelles
Brussels, Belgium

Sabine Vanhuysse
ANAGEO - DGES
Université Libre de Bruxelles
Brussels, Belgium

Eléonore Wolff
ANAGEO - DGES
Université Libre de Bruxelles
Brussels, Belgium

***Abstract*—This research aims at assessing the potential of an semi-unsupervised approach to create labeled samples for predicting deprivation probability at 100x100m. We compare Machine Learning (ML) and Deep Learning approaches (DL) with a combination of Sentinel-1 (S1) and Sentinel-2 (S2) data. Our results confirm that the propose approach for creating labelled samples using semi-supervised method works. The best performance is achieve by the ML approach combining S2 and S1 data and reach an overall accuracy of 95.34%.**

***Index Terms*—slum, deprivation, Nairobi, machine learning, deep learning**

I. Introduction

Around 1 billion people or a third of the global urban population are currently living in informal settlements and 23% of whom are situated in Sub-Saharan Africa ([1], [2]). By 2030, the United Nations estimated that this population will continue growing and probably reach 3 billion *loc. cit.*. This phenomenon is mainly caused by the fast urbanisation and the demographic growth that African countries are facing. In the frame of humanitarian aid or social concerns, it is essential to determine and monitor the spatial extent of slums in sub-Saharan African cities. The Earth observation methods seem to be the rare techniques that could achieve this task with limited time and cost. Some research integrate earth observation methods with ancillary data such as OpenStreetMap [3] but the vast majority are mostly based on the morphological and spectral features to determine the slums. The morphology of these slums is generally characterised by irregular patterns of small buildings and a scarcity of green areas [4] but in some examples the slum areas can be high-rise building with a regular pattern. Consequently, the methods being based on physical features could represent a source of limitations due to the large variety of slums.

The use of artificial intelligence and Sentinels-1/2 (S1/S2) data is a well developed and a common approach to map slums. Despite its low spatial resolution, the main benefit of S1/S2 data is to be available at no cost which allows a regular update of this fast-changing environment. A previous research by Vanhuysse, Georganos, Kuffer, *et al.* [4] used S1/S2 to generate a gridded deprivation probability over the city of Nairobi (Kenya) [4]. Beside S1/S2 images, spectral indices and textures were computed with several kernel sizes. The area of interest had been classified into 8 different classes where two of which were labelled as deprived areas. Their results showed that a supervised machine learning (ML) classifier (Random Forest) achieve 80% overall accuracy when based only on the S1 and 91% when based on S2 images, depending on the class of slum. Another study by Ienco, Interdonato, Gaetano, *et al.* [5] used Sentinel data in a deep learning (DL) approach to predict land use of the Réunion Island and Koumbia in Burkina Faso. They used a double stream model called TWINNS and inspired by the VGG model. Each branch (stream) of the model uses as input either the S1 dual-pol (VV, VH) or radiometric indices and time series from the S2. The training patches with a 5x5 pixels size have been labelled into 8 and 13 classes respectively for Koumbia and the Réunion Island. The results of the TWINNS model have been compared with multiple DL and ML architectures. Their model did not obtain the best accuracy depending on the area of interest. However, when the classes are analysed separately, the F-measurements of the TWINNS model are globally better than the metrics of the other models. Generally, while supervised approaches such as ML or DL models work well, they rely on the existence of labelled training data which is often not available and/or difficult to collect. In this paper, we investigate a hybrid unsupervised/supervised workflow to predict a gridded deprivation probability over the city of Nairobi (Kenya). More precisely, we try to assess how a unsupervised approach can be leverage to create (weakly-)labelled data to train supervised ML and DL approaches.

II. Study area and data

A. Study area

We focus on Nairobi City, the Kenyan capital, and its surroundings. The size of the area of interest (AOI) is around 664 km^2 that corresponds to a surface 6 times bigger than the city of Paris. This region is situated near the equator thus the intertropical convergent zone. As a result, our AOI is affected by a nearly constant cloud cover that limits the use of images from optical sensors mounted on satellites. The AOI has been delimited from the administrative boundaries and the morphological boundaries. The last boundaries mainly rely on the Global Human Settlement Layer Urban Centre Database that determines the degree of urbanisation ([4], [6]).

This research was partly founded by the Fund for Scientific Research (FNRS).

Post-print

The analysed region is highly heterogeneous in terms of land use. Consequently, the model will deal with a large variety of slums and other housing types (high-rise, urban, rural, ...) and green areas.

B. Data

Our study uses S1 and S2 images as well as Google Open Buildings dataset in order to determine the probability of deprivation at 1 ha. The S2 data consists of an atmospherically-corrected cloud-free temporal composite of the first quarter of 2019 to bypass the constant cloud cover typical of the tropical region and resampled at 10 meters. S1 data used for our experiments are dual-pol (VV+VH) Interferometric Wide swath mode images. Ten images have been averaged using SNAP [7] on the same period as the S2 images to lower the SAR data noise. Before averaging, the S1 have been pre-processed with thermal noise removal and terrain correction. We also use an interferometric coherence (VV) computed from 6 pairs of images with 12-days intervals between the two images (see Vanhuysse [8] for the details of the processing). Using coherence in addition to intensity was considered potentially beneficial, as high phase stability between different SAR images is expected for man-made structures such as buildings and paved surfaces. All dual-pol and coherence bands have been co-registered and resampled to match with the S2 images. The data over the whole AOI have been tiled based on a regular grid of 100x100m and normalized based on the min-max method before being used. The Google Open Buildings dataset (GOB) is used for generating the training set using a semi-unsupervised approach.

III. Methods

Our workflow is presented in *Fig. 1*. The first step consists in using GOB dataset to create labelled samples using a semi-unsupervised approach. This labelled dataset is then used to train both ML and DL supervised approaches.

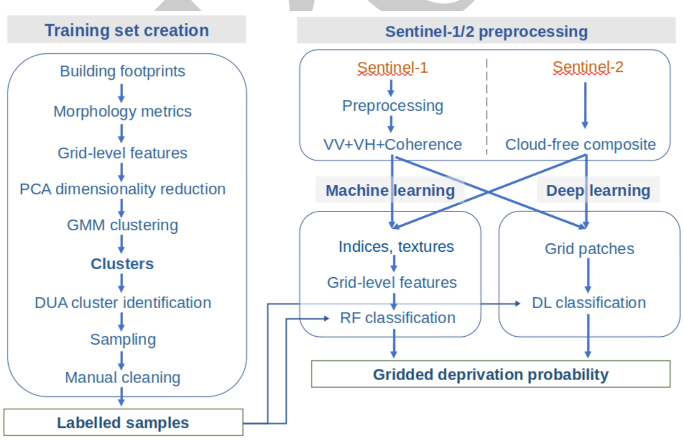

Fig. 1. *The workflow of this research.*

A. Generation of training data

In order to create labelled training data we computed 23 interpretable urban morphometrics from the GOB footprints (e.g., distance between buildings, spatial arrangements). Next, we derived features by computing four contextual statistics in grid cells (100m x 100m) and performed dimensionality reduction with principal component analysis based on a threshold of cumulative percentage of variance explained. Those features are used to group similar grid cells into 10 clusters using a probabilistic algorithm (namely Gaussian mixture). The number of clusters was determined empirically to be sure to capture Deprived Urban Areas (DUAs) and the membership associated to the cluster that best matched the spatial patterns of DUAs was used as a proxy for the morphological deprivation probability and mapped as a first output. We compared this output to DUAs outlines produced through visual interpretation of very-high resolution (VHR) WorldView-3 imagery (30 cm). After a manual cleaning to discard mislabeled or uncertain grid cells belonging to DUAs clusters, we performed a stratified sampling to select a balanced labelled training set with more or less half of them being slums and the other half labelled as non-slums.

B. ML approach

For the ML approach, we computed 20 spectral indices and textures for S2, along with S1 VV intensity, VH intensity, coherence and 18 related textures. At the grid cell level, seven statistics were computed for all indices and textures, giving a total of 140 S2-features and 147 S1-features. We used, on the one hand, only S2 features and, on the other hand, both S1 and S2 features in Random Forest (RF) models.

C. DL approach

For the DL approach, we conducted several experiments based on different S1/S2 bands combination as reported in *Tab. I*. For S2, we first use a simple combination of visible and near-infrared bands (RGBNIR). We also design two experiments with SWIR bands (BNIR and RNIR) because previous studies determined that the red and the short-wave infrared bands (SWIR) have the biggest impact on the slum prediction ([4], [9]).

For S1, the 'ALLS1' combination include dual-pol images (VV and VH) and the coherence create by [4]. To determine whether the coherence band is important, the 'VVVHS1' combination has been built only with the dual-pol S1 bands.

We design the YMCA model (Y-Model for Classification by ANAGEO) (*fig. 2*) to predict deprivation probability. This is a Y-shape functional model mainly inspired by Atienza [10] (pp. 47-53) and modified to fit with our patch size, improving the validation accuracy and reducing overfitting.

The model has two separate inputs (10x10x4 for S2 ; 10x10x3 for S1) which are concatenated to apply the same data augmentation (random flip horizontally and vertically and random rotation with max angle of $\pm 72°$). Then, the S1/S2 augmented layers are separated into two parallel branches and processed independently. The images follow two sequences

Post-print

TABLE I
Combination of the Sentinel-1/2 level-1 images for the different experiments.

	Sentinel-2			Sentinel-1	
Bands	BNIR	RNIR	RGBNIR	ALLS1	VVVH
1	B02 (Blue)	B04 (Red)	B04 (Red)	Coherence	VV
2	B08 (NIR)	B08 (NIR)	B03 (Green)	VV	VH
3	B8A (NIR)	B8A (NIR)	B02 (Blue)	VH	
4	B12 (SWIR)	B12 (SWIR)	B08 (NIR)		

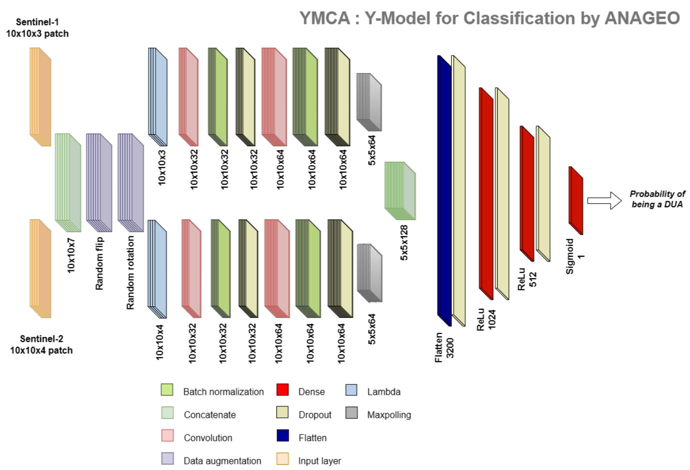

Fig. 2. *Representation of the YMCA model.*

of 3 layers: a 2-dimensional convolution layer with a kernel of 2x2, a batch normalisation layer and a dropout layer. Finally, each branch ends with a maxpooling layer. Then, both branches are concatenated and flattened into a one-dimensional vector followed by three dense+dropout layers. The output layer has only one neuron with a sigmoid activation function that attributes a score of being a DUA. In other words, the model outputs value between 0 and 1 and patches with values above 0.5 are considered as DUAs. For each k-fold (see after), the weights of the best epoch are saved and used to predict the entire scenery.

D. *Performance evaluation*

For training and validation purposes, we use 6922 tiles derived from an unsupervised clustering performed on the same statistics as those used in Vanhuysse [8] and by photointerpreting the clusters belonging to the deprived areas. Performance was evaluated through both 5-fold cross-validation (80/20) and we compare the predictions of the models to the manually delineated DUA outlines created by a local expert.

IV. Results

The ML approach achieves an overall accuracy of 93.7% when using only S2 features, and reaches 95.6% when used together with S1 features. Similarly, the DL results show that the combined use of BNIR and ALLS1 bands achieve the best accuracy ($\pm$ 89%) compared to the other experiments. The impact of discarding S1 in the DL experiments was variable depending on the S2 bands used, with losses between 1% and 5%. When BNIR is replaced by RNIR for the S2 data, the k-fold standard deviation is increasing significantly and the performance is lower. Interestingly, the RGBNIR dataset combined with ALLS1 achieves similar performance to the best-performing combination BNIR+ALLS1, with less than 1 percentage point of difference. When looking on the performance of the DUA class only, F1-score achieve acceptable for the ML approach but significantly weaker for the DL one.

The performance metrics computed on the independent test set confirm the results of the kfold cross-validation, with the ML approach outperforming the DL one, reaching respectively 95.34% and 93.05% overall accuracy. While the difference in terms of overall accuracy is limited, the visual inspection of cartographic results, as visible on *Fig. 3*, shows clear differences. The DL approach under-predicts for slums, especially in the north of the city and over-predicts a lot in the Western and southern part of the city. The better performance of the ML approach could be explained by the use of texture rasters with kernel sizes bigger than the patch size used in DL. Indeed, some textures uses a kernel of 11x11 pixels and could extract patterns in a larger depth of field, better suited to discriminate DUAs and non-DUAs.

V. Conclusion

Our research demonstrates that semi-unsupervised approach combining local-expertise and visual interpretation could be used to generate labelled samples for predicting deprivation probability on 100x100m gridded products. Also, we show the added-value of combining S1 in addition to S2 data, either using ML or DL approach.

Acknowledgment

We warmly thank Kabiru Priscilla for her participation in the visual interpretation and hand-made delineation of DUAs outlines.

References

[1] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," en, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 59–69, Apr. 2019, ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2019.02.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271619300383 (visited on 07/28/2022).

Post-print

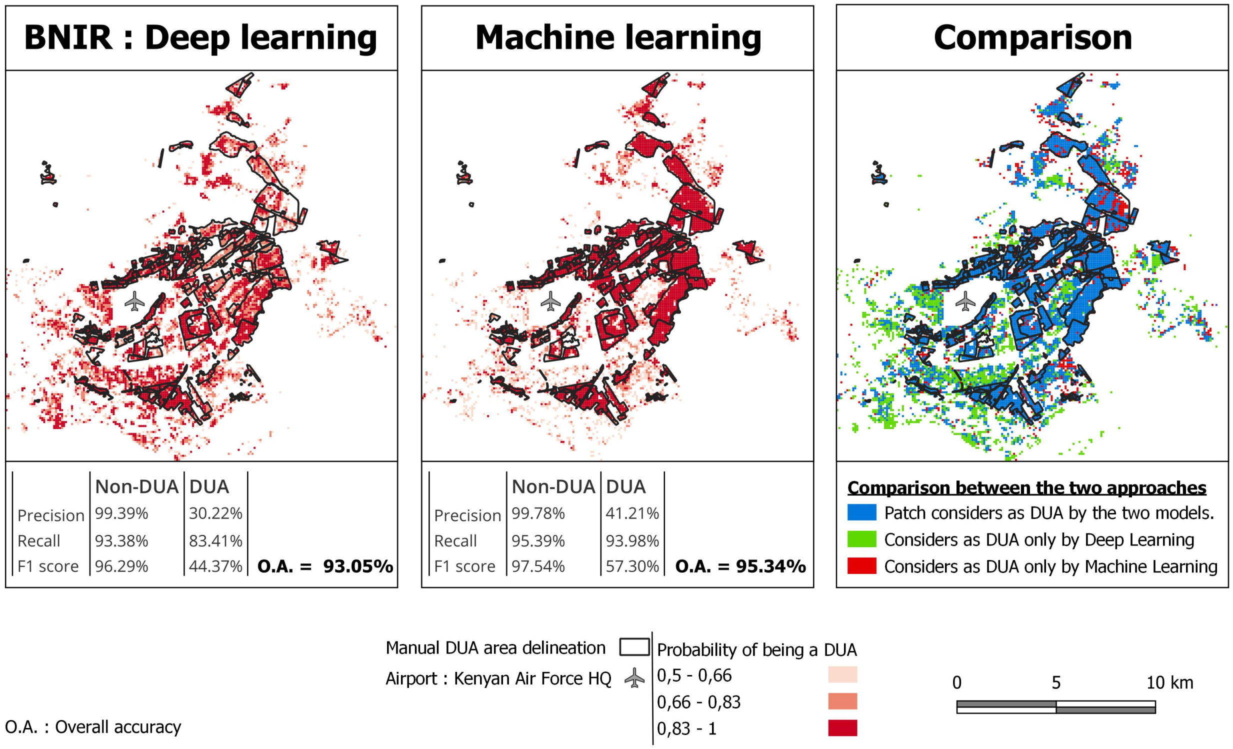

Fig. 3. *Comparison of DL and ML prediction maps.*

[2] United Nations : Department of Economic and Social Affairs, *Goal 11 : Sustainable cities and communities*, 2019. [Online]. Available: https://unstats.un.org/sdgs/report/2019/goal-11/ (visited on 08/08/2022).

[3] B. Ayo, "Integrating openstreetmap data and sentinel-2 Imagery for classifying and monitoring informal settlements," eng, Feb. 2020, Accepted: 2020-03-02T17:14:07Z. [Online]. Available: https://run.unl.pt/handle/10362/93641 (visited on 07/28/2022).

[4] S. Vanhuysse, S. Georganos, M. Kuffer, T. Grippa, M. Lennert, and E. Wolff, "Gridded Urban Deprivation Probability from Open Optical Imagery and Dual-Pol Sar Data," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, ISSN: 2153-7003, Jul. 2021, pp. 2110–2113. DOI: 10.1109/IGARSS47720.2021.9554231.

[5] D. Ienco, R. Interdonato, R. Gaetano, and D. Ho Tong Minh, "Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture," en, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 11–22, Dec. 2019, ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2019.09.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271619302278 (visited on 08/03/2022).

[6] A. Florczyk, M. Melchiorri, C. Corbane, *et al.*, *Description of the GHS Urban Centre Database 2015*. Jan. 2019. DOI: 10.2760/037310.

[7] European Space Agency, *SNAP (ESA's SentiNel Application Platform)*, 2022. [Online]. Available: https://step.esa.int/main/download/snap-download/ (visited on 08/22/2022).

[8] S. Vanhuysse, *Description of sentinel-1 preprocessing*, 2022. [Online]. Available: https://github.com/jgovoort/slumap-ymca/blob/main/S1_pre_processing/S1_preprocessing_main_Vanhuysse_2022.pdf (visited on 08/23/2022).

[9] T. Fisher, H. Gibson, Y. Liu, *et al.*, "Uncertainty-Aware Interpretable Deep Learning for Slum Mapping and Monitoring," en, *Remote Sensing*, vol. 14, no. 13, p. 3072, Jan. 2022, Number: 13 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-4292. DOI: 10.3390/rs14133072. [Online]. Available: https://www.mdpi.com/2072-4292/14/13/3072 (visited on 07/26/2022).

[10] R. Atienza, *Advanced deep learning with TensorFlow 2 and Keras: apply DL, GANs, VAEs, deep RL, unsupervised learning, object detection and segmentation, and more*, eng, Second edition. Birmingham Mumbai: Packt, 2020, ISBN: 978-1-83882-165-4.

Post-P